

OXFORD RECORD LINKAGE STUDY

A CENTRAL FILE OF MORBIDITY AND MORTALITY RECORDS FOR A PILOT POPULATION

BY

E. D. ACHESON

Nuffield Department of Clinical Medicine, University of Oxford

The application of medical records to serve only their own section of the Health Service is outdated—the individual is the unit and not the section of the Health Service.

P. Bothwell (1961).

The Chief Medical Officer of Health for Scotland has recently reviewed the various types of medical data collected routinely from which information concerning health, sickness, and mortality in Britain can be derived (Cowan, 1960). All these data (Registrar General, 1921–60, 1954–56; Ministry of Health, 1958; Ministry of Pensions, 1958–9; Logan, 1953; Logan and Brooke, 1957; General Register Office, 1956) produce information concerning the national health in cross-section in terms of certain defined events; they permit valuable estimates to be made of the number of consultations, admissions, discharges, claims, or deaths per unit population; but their disadvantage is that they tell little or nothing concerning the combinations of events implicit in the terms prognosis, differential survival, recurrence, transfer, and readmission, nor about the association (positive or negative) between one disease and another: and nothing whatever concerning the patterns of disease in families.

The National Cancer Register (Registrar General, 1957) involves a different principle, that of bringing together information concerning subsequent events to the same patient into a single file, thus permitting the calculation of incidence rates, and within the particular field of malignant disease making possible estimates of prognosis. For mental health, Tooth and Brooke (1961) have tackled the problem of prognosis by defining cohorts based on all patients first admitted to mental hospitals in 1954, 1955, and 1956, and by following their course in and out of mental hospitals. In both these instances valuable longitudinal data are accruing concerning certain limited aspects of the health of individuals.

The question which is currently being examined by the Oxford Record Linkage Study is whether, as a result of the revolution in data-processing techniques, it is now practicable to provide a single integrated file of health data for a community in which the main events occurring to persons will not only be recorded, but will also be brought together in such a way as to permit both cross-sectional and longitudinal patterns of events to be made out, and familial aggregations of disease studied.

THE IDEA OF RECORD LINKAGE

The bringing together of separately recorded data concerning the important health events in the life of individuals for a whole population was first envisaged by William Farr (1874). He recognized the importance of data which would permit both cross-sectional and longitudinal studies. However, in the absence of the practical means for its accomplishment, record linkage had still not been carried out on any scale when Stocks (1944) and Dunn (1946) reiterated its advantages 70 years later. It remained for Canada to take the first step in such a process by preparing a National Index of birth, marriage, and death certificates (Marshall, 1947), which consists of an alphabetical listing of all the vital events for the whole country derived mechanically from cards punched from microfilm copies of the certificates. In recent years Newcombe, James, and Axford (1957) and Newcombe, Kennedy, Axford, and James (1959) have made two further important methodological contributions:

(a) They have linked stillbirth and birth certificates with marriage certificates, thus forming a nucleus of family groupings of records for the population of British Columbia.

(b) They have taken advantage of the recent technical revolution in methods of data-processing

and have shown that, given the liberal amount of identifying data currently recorded on Canadian vital certificates, it is possible to programme a computer to link them into family groups speedily and economically, in spite of discrepancies in individual items of identification.

In this way the vital certificates are being used to trace out the lattice-work of blood relationships in a community. This in itself is a fundamental step: however, until this matrix can be clothed with medical information recorded at birth, school medical examination, discharge from hospital, and death, the accumulation of material useful for medical research is bound to be slow. Were it possible to introduce into the system the main diagnoses reached in the course of a lifetime, light would be cast on many of the problems not only of the natural history of disease but of human genetics. A similar national register (but excluding hospital data) has been proposed for Belgium by Twiesselmann (1962). Bothwell (1960) has gone so far as to say that "the parochial view of records as hospital, local authority, general practice, social service, or district entities with their unrelated functions is outmoded and a severe handicap to progress in any field of human care". The implications of record linkage in population genetics and epidemiology have recently been discussed in a symposium by Cavalli-Sforza, Doll, Dunn, Binder, and others (United Nations, 1960).

OXFORD RECORD LINKAGE STUDY (ORLS)

The opportunity afforded by the National Health Service for the development of a central health register in which morbidity data would be linked into a framework of information derived from the records of vital events was recently outlined by Acheson, Truelove, and Witts (1961). Shortly afterwards, thanks to a grant from the Nuffield Foundation, a pilot study of Record Linkage was set up in the Oxford area. The minimum requirements of the method are these:

- (1) Individuals within the file will be identifiable beyond reasonable doubt;
- (2) There will be sufficient data to place an individual in his family;
- (3) The main health events of the individual's life will be recorded;
- (4) Information will be available to enable the source documents to be found, *e.g.*, in the case of hospital data, the hospital number;

- (5) Sufficient administrative data will be recorded to permit services to be offered to co-operating hospitals, health departments, etc., which will help to pay the expenses of the scheme. What other data should be stored depends upon its utility, availability, reliability, and the additional expense involved.

The main methodological problems are thus the cheap acquisition of data, the identification of medical records, and the collation, storage, and retrieval of the data by modern methods.

(a) THE ACQUISITION OF DATA.—The pilot population consists of the residents of the County Borough of Oxford, Oxfordshire (except Henley M.B. and R.D.C.), and of Abingdon Borough and Abingdon Rural Area. It comprises a total of 325,000 persons, subject to increase from births and immigration, and to decrease from deaths and emigration. Data have been collected concerning all births, all confinements (whether at home or in hospital), all hospital discharges (including mental and long-stay), and all deaths which have occurred in this population since January 1, 1962 (Table I). This has involved collating selected data from 101 files. Information concerning births and deaths is transcribed from the appropriate notifications in the departments of the Medical Officers of Health. Data concerning discharges from hospital are transcribed by ORLS clerical staff in the 29 Hospital Record Departments. This expensive process accounts for a substantial part of the cost of the study.

TABLE I
EVENTS RECORDED FOR PERSONS RESIDENT IN THE
DEFINED AREA ON OR AFTER JANUARY 1, 1962

Event	Document	Source
Birth ..	1962 Birth notification 1963 Birth entry	Medical Officers of Health Registrar General
Domiciliary Confinement	Midwives' records	Domiciliary Midwives
Hospital Confinement	Hospital In-patient summary sheet	Hospital Records
Stillbirth	Stillbirth notification	Medical Officers of Health
Discharge from Hospital	Hospital In-patient summary sheet	Hospital Records
Death ..	1962 Death notification 1963 Death entry	Medical Officers of Health Registrar General

(b) THE IDENTIFICATION OF THE DOCUMENTS TO BE LINKED.—The identification data recorded must be sufficient to permit records to be brought together in spite of changes of surname on marriage or of address, and a surplus must be provided to allow for the inevitable errors. Furthermore, if the documents of different members of a family are to be brought

together, at least one key document, for example the birth certificate, must also have sufficient data to identify the parents of the individual.

The practical value of an item of identification in the linking of records depends upon the following:

- (i) *Its Discriminatory Power.*—This depends upon the frequency of occurrence of the item in the population. Thus a unique given number like the National Health Service or Ministry of Pensions and National Insurance Number would be sufficient by itself to discriminate between any two persons if it were always quoted correctly*. However, names, dates (*e.g.* date of birth or date of marriage), and other characteristics also have discriminatory power according to their frequency, and can together identify an individual or a family, although compared with a given number they are relatively cumbersome.
- (ii) *Its Liability to Change.*—This limits, but does not obviate, the value of such items as surname in women, address, and occupation.
- (iii) *Its Liability to Error.*—Errors have been found in almost all types of identifying data, commonly in forenames, birth surname, and year of birth, but less commonly in current surname and day and month of birth. Various techniques are available to minimize the effect of minor alterations in the spelling of names (Newcombe and others, 1957, 1959).
- (iv) *Its Availability on Each of Two Documents to be Linked.*—This in turn depends upon

the extent to which the item is known to the individual or, in the case of a death certificate, to the informant. The N.H.S. Number is not at present freely available (Table II).

To these four should be added a fifth, namely the expediency or desirability of obtaining the item of information from persons. Thus, date of marriage and mother's maiden name (Newcombe and others, 1957; Smith, 1962) are highly discriminatory pieces of information in the linkage of documents into family groups, but it is not expedient to ask for them routinely on admission to hospital. If the N.H.S. Number could be made available on medical documents the additional discriminatory power so provided would greatly simplify the linkage of records. However, in Great Britain, there is reluctance to carry evidence of identity in the form of a number on a disk or card, although such a card might have practical advantages to the individual if it also contained information concerning immunization against tetanus and drug idiosyncrasies (Hansard, 1963). The use of a similar number provided at birth (or immigration) is being promoted in Canada (Marshall, 1947), where its wider use was recently recommended by the Royal Commission on Government Organization (Glassco Commission, 1962), and in Belgium (Twisselmann, 1962).

HOSPITAL RECORDS.—The items of identification commonly used by hospital records officers are current surname, forenames, sex, date of birth, address, and hospital number. As these might not be sufficient in a very large file, records officers in the ORLS area were asked to co-operate by trying to obtain three additional constant items, birth surname, birthplace (town and county), and N.H.S. Number. The availability of these items in a sample of 1,202 hospital discharges in April, 1962, is shown in Table II.

TABLE II
FREQUENCY WITH WHICH VARIOUS ITEMS OF IDENTIFICATION WERE RECORDED ON A SAMPLE OF HOSPITAL IN-PATIENT SUMMARY SHEETS IN APRIL, 1962, 3 MONTHS AFTER THE COMMENCEMENT OF THE STUDY

Item of Identification	Men. Single Women and Children		Married and Widowed Women		All	
	No.	Per cent.	No.	Per cent.	No.	Per cent.
Current Surname	810	100·0	392	100·0	1,202	100·0
One Forename	810	100·0	392	100·0	1,202	100·0
Date of Birth	774	95·6	364	92·9	1,138	94·7
Place of Birth	551	68·0	262	66·8	813	67·6
Maiden Name	—	—	272	69·4	—	—
N.H.S. No.	307	37·9	170	43·4	477	39·7

* In a sample of pairs of documents relating to the same individuals, discrepant versions of the N.H.S. Number were encountered in more than 10 per cent.; the proportion of discrepant Midland Personnel Numbers was roughly similar (Hogben and Cross, 1948).

BIRTH, STILLBIRTH, MARRIAGE, AND DEATH CERTIFICATES.—At the present time the identifying data on these certificates in England and Wales are neither sufficient to identify the individual over a lifetime (*i.e.* when changes in name on marriage and changes in address have occurred), nor are they adequate to permit the records easily to be grouped into families. In Table III are set out some of the additional data recorded on the entries in the relevant American, Scottish, and Canadian Registers of Vital Events. That it is practicable to obtain such information is shown by the experience of the Registrar General for Scotland (1963), who obtained the maiden name of the deceased person's mother in 99·4 per cent. of a sample of Edinburgh death certificates. If the bulk of this data were recorded in England and Wales, and age were replaced by date of birth, the linkage of vital certificates into family groups on a large scale could become a practical proposition. According to the World Health Organization "Handbook of Vital Statistics", the deceased person's date of birth is already recorded on the death certificate by 26 countries and his birthplace by 29 countries.

(c) **COLLATION, STORAGE, AND RETRIEVAL OF INFORMATION.**—The transcribed data is coded centrally, standard codes being used wherever possible (WHO, 1957; General Register Office, 1960). The process of linkage of the records involves two stages which can be performed either clerically or, as Newcombe and others (1959) have shown, by an electronic computer:

(1) The sorting of the records into sequences which will permit pairs of documents likely to match to be brought together for scrutiny;

(2) The decision whether or not the two records originate from the same individual in spite of discrepancies or omissions in the identifying particulars.

At present both these stages are carried out clerically, using a master index arranged alphabetically by maiden surname with cross-index cards under married surname for married women. The data are then coded and punched on cards. Information concerning different events which have occurred to an individual can then be brought together at will. The nucleus of a file of family records has been attained by punching a number identifying the mother on the baby's birth card. Thus the health records of mother and baby can also be brought together. In 1963, it will be possible to add some identification particulars of the father.

The analysis of the first year's data will be by standard mechanical sorting and tabulating machines.

FUTURE PLANS

The main achievement to date has been the creation of a central linked file of morbidity and mortality data for a defined population. The master index already contains some 40,000 entries, and the processing of the first year's data is almost complete. The following are examples of the types of analysis envisaged in the short term: readmission rates, interhospital transfer patterns, a study of death at home after discharge from hospital, and the relationship of certain perinatal events (for deliveries at home and in hospital) to hospital morbidity and mortality in the first year of life.

It is also hoped to provide certain services to co-operating hospitals in the form of administrative statistics and listings*. The system can also be used for studies in operational research and quality control such as those successfully carried out in North America by the Professional Activity Study (1961). In this way the study may eventually pay its way.

* Including lists of deceased persons to permit clearance of current files of records.

TABLE III

SOME OF THE DATA RECORDED IN THE REGISTERS OF VITAL EVENTS IN SCOTLAND (S), CANADA (C), AND THE UNITED STATES OF AMERICA (A), BUT NOT IN ENGLAND AND WALES

Birth and Stillbirth Entries		Marriage Entry		Death Entry	
A.C.	Mother's place of birth	A.C.	Bride's place of birth	A.S.C.	Maiden name of deceased person
A.C.	Father's place of birth	A.C.	Groom's place of birth	C.	Date of birth of deceased person
S.	Place of marriage	S.C.	Bride and Groom's mothers' maiden names	A.C.	Place of birth of deceased person
S.	Date of marriage*	S.C.	Bride and groom's mothers' given names	A.S.C.	Names (incl. mother's maiden name) of deceased person's parents
		A.	Bride's date of birth		
		A.	Groom's date of birth		

* Recorded among the confidential particulars on English birth certificates.

It will also be possible to produce a limited amount of material bearing on the epidemiology of diseases for which hospital admission is the rule. The long-term results of a more extended file have already been set out elsewhere (Acheson and others, 1961).

In future, efforts will also be directed towards cutting out some of the manual methods at present in use as these are expensive, time-consuming, and subject to human error. Eventually it is hoped to abolish the manual transcription of data by introducing a degree of standardization of in-patient summary sheets so that coding can be done directly from photo copies (Bothwell, 1960), to transfer the index on to magnetic tape, and to programme a computer to perform the linkage process.

One of the most difficult problems concerns the conflicting claims of depth of coverage and the size of the population to be covered. The size of the population is important because it determines how quickly material will accumulate concerning an event or pattern of events of a given frequency. With larger populations it will be possible to study rarer events more quickly. The ideal size of population is also related to the problem of emigration. Other things being equal, the smaller the population the more wastage there will be through emigration and the more attenuation of longitudinal data.

There is no doubt that in the current study the population is too small. If it is to be extended substantially it will be necessary to revise the vital certificates for England and Wales so that they contain adequate data to identify the individual in spite of changes of surname on marriage and of address, and to permit identification of the parents on at least one certificate. For the economic handling of very large files it is probable that the additional discriminatory power of some sort of "given number", like the N.H.S. Number, will be necessary on the incoming "amending" items.

That modern equipment permits the handling of very large files of data indeed is shown by the experience of the Ministry of Pensions and National Insurance with the National Graduated Pensions Scheme. By means of a large electronic data-processing installation, it has been found possible to update a file of 27,500,000 accounts at the rate of 350,000 amending items a day and to answer 7,000 inquiries every second day. Approximately four-fifths of the pairs of documents are linked automatically by the computer on the basis of MPNI number, first three characters of surname, first initial, the matching of the remaining 20 per cent. of discrepant and deficient documents being accomplished clerically. A National Register of births, marriages, deaths, and discharges from hospital for

England and Wales with 100 per cent. coverage would involve less than 20,000 amendments a day in a file of 45,000,000 records. However, the data would be less uniform and the burden of coding would be heavier. The clerical basis for such a system is already in existence in the Central Register and Index of the National Health Service at Southport.

The current pilot population also requires further coverage in depth. The diagnostic information is obviously very far from complete. A modest addition to the data would be an abstract of the positive findings at the first school health examination and from the records of special schools. The addition of hospital out-patient data or material from family doctors would be a much larger undertaking.

SUMMARY

The framework of an integrated file of morbidity and mortality data for a population of 325,000 persons is now in existence. Within this file the working units are the individual and the family. Data are recorded for the following events: birth, domiciliary and hospital confinement, stillbirth, discharge from hospital, and death. Where more than one of these events has occurred to an individual since January 1, 1962, the records concerning them have been brought together. In addition, the health records of babies born since January 1, 1962, have been linked with those of their mothers, thus forming the nucleus of a file of family records. The processing of the first year's data is almost complete. Future efforts will be directed to abolishing as many manual processes as possible and to analysis.

At this early stage it is possible to discern certain changes that would be necessary if this type of integrated file of health documents were to be extended. The amount of identifying data on birth, stillbirth, marriage, and death certificates in England and Wales would need to be increased; the amount of identifying data on the records of co-operating hospitals would also need to be amplified and in-patient summary sheets designed according to certain standard principles. A wider use on medical documents of a N.H.S. Number would simplify the document-matching procedures.

The Oxford Record Linkage Study is supported by a generous grant from the Nuffield Foundation, and has been provided with a Headquarters by the Oxford Regional Hospital Board. Its existence results from the co-operation of many persons, of whom I wish particularly to acknowledge Mr. A. Barr, the Chief Statistician and Records Officer of the Oxford Regional Hospital Board, and his Staff, the Medical Officers of Health, and

the Records Officers of the Hospitals in the Area. I should also like to thank Prof. L. J. Witts, C.B.E., and Dr. S. C. Truelove for reading the typescript and making constructive criticisms.

REFERENCES

- Acheson, E. D., Truelove, S. C., and Witts, L. J. (1961). *Brit. med. J.*, **1**, 668.
- Bothwell, P. W. (1960). *Med. Rec.*, **5**, 298, 320.
- (1961). *Ibid.*, **5**, 359.
- Cowan, K. (1960). In "Proc. III Int. Congr. Medical Records, Edinburgh, 1960", p. 15.
- Dunn, H. L. (1946). *Amer. J. publ. Hlth*, **36**, 1412.
- Farr, W. (1875). "Supplement to 35th Annual Report of the Registrar General", p. 110.
- General Register Office (1960). "Classification of Occupations, 1960". H.M.S.O., London.
- (1956). "General Practitioners Records, 1952-54". Studies on Medical and Population Subjects No. 9, in continuation of No. 7. H.M.S.O., London.
- Hansard (1963). "Parliamentary Debates, House of Commons", 5th series, vol. 673, col. 946.
- Hogben, L., and Cross, K. W. (1958). *Brit. J. soc. Med.*, **2**, 149.
- Logan, W. P. D. (1953). "General Practitioners Records". Studies on Medical and Population Subjects, No. 7. H.M.S.O., London.
- and Brooke, E. M. (1957). "The Survey of Sickness, 1943-52". *Idem*, No. 12.
- Marshall, J. T. (1947). *Pop. Stud.*, **1**, 204.
- Ministry of Health (1960-61). "Report on Hospital In-Patient Enquiry for 1958". H.M.S.O., London.
- Ministry of Pensions (1958-59). "Digest of Statistics analysing Certificates of Incapacity". (Not for publication).
- Newcombe, H. B., James, A. P., and Axford, S. J. (1957). "Family Linkage of Vital and Health Records". Atomic Energy of Canada, Report No. 470.
- , Kennedy, J. M., Axford, S. J., and James, A. P. (1959). *Science*, **130**, 954.
- Registrar General for England and Wales (1921-60). "Statistical Reviews of England and Wales".
- (1954-6). *Idem*. Supplement on Mental Health.
- (1957). *Idem*. Supplement on Cancer, 1952.
- Registrar General for Scotland (1963) Personal communication.
- Royal Commission on Government Organization (Glassco) (1962). Volume 1, "Management of the Public Service". Queen's Printer, Ottawa.
- Smith, A. (1962). Personal communication.
- Stocks, P. (1944). *Proc. roy. Soc. Med.*, **37**, 593.
- Tooth, G. C., and Brooke, E. M. (1961). *Lancet*, **1**, 710.
- Twisselmann, F. (1962). *Bull. Acad. roy. Méd. Belg.* 7th series, **2**, 93.
- W.H.O. (1957). "Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death".
- (n.d.). "Handbook of Vital Statistics Methods", Statistical Papers, Ser. F, No. 7. (U.N. Publication, Sales No. 55, XVII, 1).
- (1962). "Use of Vital and Health Statistics for Genetic and Radiation Studies".